CLS Access to Data not Available via the UK Data Service

[1. Summary 1](#_Toc526694224)

[2. CLS Data Access Application Form 2](#_Toc526694225)

[3. Guidelines to access data from the CLS cohort studies 6](#_Toc526694226)

[4. CLS Data Classification Scheme for Data Sharing 8](#_Toc526694227)

# Summary

The Centre for Longitudinal Studies (CLS) welcomes proposals to access CLS data in the exceptional circumstance that they are *not currently available* via the defined mechanisms governing the release of CLS data, which are the UK Data Service (UKDS) for the main survey data, and METADAC or WTCCC DAC for genotypic data. The detailed [guidelines for data access](http://stage.ninjasforhire.co.za/cls/wp-content/uploads/2017/06/Data-Access-Guidelines-CLS-Data-Access-Committee.pdf) can be found in section 3 of this document.

Data access applications may include (but are not restricted to):

* **Main survey data**: some data collected by the CLS cohort studies have not been deposited at the UKDS due to their sensitive nature or because they have not been prepared sufficiently for deposit.
* **Linked administrative records:** CLS has a programme of record linkages underway and the data are made available through the UKDS. However, some of these linked data have not been deposited at the UKDS due to their sensitive nature, but are available for research purposes under strict Secure Access conditions from the CLS premises.
* **Paradata:** CLS holds data about the data collection process from some data collection sweeps. These are collected primarily for administrative purposes and are not routinely released for research use.

If you wish to access CLS data not currently available via the UK Data Service (or another mechanism) you should first contact [clsfeedback@ucl.ac.uk](mailto:clsfeedback@ucl.ac.uk) to explore the likely feasibility of the request. Applicants should next complete the following application form and send it to [clsfeedback@ucl.ac.uk](mailto:clsfeedback@ucl.ac.uk). All proposals will be discussed at a meeting of the **CLS Data Access Committee** (CLS DAC), which typically meets every month.

Researchers must allow plenty of time between submitting their application and when they plan to undertake research on the data which they are applying for. In cases where significant resources of the CLS data management team are required to fulfil the request, we suggest an application is submitted at least 3 months before the planned research is due to take place.

Please note that the CLS DAC also considers proposals for data enhancements and external record linkages that have not yet been enacted. These proposals are governed by separate guidelines and application procedures.

# CLS Data Access Application Form

*DAC REF:*

## Main data applicant

For PhD projects, the main data applicant on the application must be the supervisor.

|  |  |
| --- | --- |
| **Title, forename, surname:** |  |
| **Work Address:** |  |
| **Academic email address:** |  |
| **Telephone:** |  |

## Research team

Please list the names of all member(s) of your research team who need access to the data and their contact email address.

|  |  |
| --- | --- |
| **Name** | **Email address** |
|  |  |
|  |  |
|  |  |
|  |  |

## Location

Please specify the location where the data will be accessed and stored if it is different to the organisation/institution specified in section 1.

|  |
| --- |
|  |

## Type of data

Please indicate what not deposited data are being requested:

Main survey data

Transcribed or digitised legacy data

Linked administrative records

Paradata

Other (please specify)

## CLS Cohort(s)

Please indicate which CLS study or studies this proposal relates to:

National Child Development Study (1958 Birth Cohort)

1970 British Cohort Study

Next Steps

Millennium Cohort Study

## Title of the research project

## Abstract of the research project

Please provide a short description of the proposed research you plan to do with the requested data, including main research question(s):

|  |
| --- |
|  |

## Details of the research proposal

Please provide a full and detailed description of the purpose for which the data are requested, describing the aims of the study/research and the potential impact of the research for policymakers and/or practitioners. Where research is part of a larger programme, please include details below.

|  |
| --- |
|  |

## Data requested

Please specify what CLS data to which access is requested. Please provide as detailed a description as possible. If applicable, please list all survey variables, or external data which you require to be linked to the data you are requesting:

## Derived variables

Please describe what derived variables or other data outputs are planned based on your proposed access to this data. All derived variables and related documentation must be returned to CLS as a condition of access.

## End date of the project

Please specify when you estimate that the project will finish (Month/Year)

## CLS conditions of data use

**All data from the CLS cohort studies is documented and archived for potential use by the wider research community.** Therefore, I agree that should access be granted, I will agree to use the data according to the CLS conditions of data use listed below. I understand that the data that I request, and any returned variables, will be documented and deposited at the UK Data Archive, and may be shared by CLS for wider use.

The CLS conditions of data user are:

* I will use the data only for the purposes of research specified in my DAC application
* I will return all derived variables and related documentation arising from this request to CLS within 12 months of receipt of the requested data.
* I will register at the UK Data Archive (<https://www.ukdataservice.ac.uk/get-data/how-to-access/registration/uk-he-fe-users>) and will abide by the terms of UK Data Archive Licence which I have signed as part of my registration.
* I will abide by the UKDS Microdata Handling and Security: Guide to Good Practice <https://www.ukdataservice.ac.uk/media/604725/cd171-microdatahandling.pdf>
* I will not link between the research identifiers supplied by the UK Data Service and any other identifiers previously issued.
* I will notify CLS of any non-compliance I am aware of
* I will not use the data for commercial purposes
* I understand that CLS does not transfer any interest in intellectual property to me
* I understand that the materials are provided without warranty or liability of any kind
* I will abide by any further conditions notified to me
* I will ensure that the means of access to the data (such as passwords) are kept secure and not disclosed to anyone else
* I will preserve the confidentiality of, and not attempt to identify, individuals, households or organisations in the data
* I will use the correct methods of citation and acknowledgement in publications
* I will send CLS bibliographic details of any published work based on the data
* I will notify CLS of any errors discovered in the data collections
* I will store the data following the guidelines data specified the UKDS Microdata Handling and Security: Guide to Good Practice
* At the end of the project, I will destroy the data as specified in section 6 of the UKDS Microdata Handling and Security guide
* I understand that any non-compliance with these terms and conditions will lead to immediate termination of my access to the data could result in legal action against me.

**Signed: Print Name:**

**Date:**

# Guidelines to access data from the CLS cohort studies

These guidelines apply to data that are not available via the UK Data Service (UKDS) for the following studies:

* National Child Development Study (NCDS, 1958 British Birth Cohort)
* 1970 British Cohort Study (BCS70)
* Next Steps
* Millennium Cohort Study (MCS)

These guidelines are derived from the CLS Data Access Framework that sets out the principles and procedures by which all access to and use of CLS data is governed.

## Introduction

### The Centre for Longitudinal Studies (CLS) is responsible for four national cohort studies: the National Child Development Study (NCDS, or the 1958 Birth Cohort Study), the 1970 British Cohort Study (BCS70), the Millennium Cohort Study (MCS) and Next Steps, which follow the same group of people from childhood and throughout their lives. CLS data is made available for researchers primarily through the UK Data Service (UKDS), although there are separate access mechanisms for the use of biological samples and genetic data arising from such samples. Section 4 of this document further describes the main mechanisms by which the majority of CLS data is made available.

### These guidelines set out how researchers may access data from CLS studies in the exceptional circumstances where these are not available via UKDS, or the other defined mechanisms governing the release of biological data.

### In exceptional circumstances where data arising from CLS studies is not available via UKDS (or alternative mechanisms for specialised biological data), applications will be considered by the CLS Data Access Committee. Such applications may include (but are not restricted to):

* + **Main survey data**: some data collected by the CLS cohort studies have not been deposited at the UKDS due to their sensitive nature or because they have not been sufficiently processed to be deposited.
* **Linked administrative records:** CLS has a programme of record linkages underway and the data are made available through the UKDS. However, some of these linked data have not been deposited at the UKDS due to their sensitive nature, but are available for research purposes under strict secure conditions from the CLS premises.

**Paradata:** CLS holds data about the data collection process from some data collection sweeps. These are collected primarily for administrative purposes and are not routinely released for research use.

Additionally the CLS data access committee also considers applications for data enhancements, that is new data collections, or external record linkages that have not yet been enacted. Applications for data enhancements are governed by separate guidelines.

## Process for requesting access to CLS data not deposited at the UKDS

### Those wishing to access any data arising from one of the CLS cohort studies should first check whether this data is available via the UK Data Service (see hyperlinks in Section 4)

### Those wishing to access data that is not currently available via the UK Data Service should first contact [clsfeedback@ucl.ac.uk](mailto:clsfeedback@ucl.ac.uk) to explore the likely feasibility of the request.

### The next step is to complete an application form. All proposals will be discussed at a meeting of the CLS Data Access Committee (CLS DAC), which meets typically every month.

### Researchers must allow plenty of time between submitting their application and when they plan to undertake research on the data which they are applying for. In cases where significant resources of the CLS data management team are required to fulfil the request, we suggest an application is submitted at least 3 months before the planned research will take place.

### Decisions about which requests will be approved will be based on the information provided in the application form and a clear set of criteria as follows:

* Risk of disclosure in respect of one or more study respondents;
* Sensitivity of data requested;
* Welfare of one or more study respondents;
* Impact of the conduct or reporting of the research on public perception or the future viability of the data collection;
* Data security;
* Where relevant, sample depletion (note, depletable biological samples fall within the remit of the METADAC);
* Public benefit, including potential scientific and wider impacts of the proposed research;
* Other possible ethical issues;
* The CLS resources required to deliver the enhancement, and any risk posed by the fulfilment of the request to CLS’ ability to deliver on its core mission or other existing commitments.

### Following discussion at the CLS DAC meeting, the data requester may be asked to respond to any queries raised by the meeting and to modify their proposal to render it suitable for further consideration.

### All researchers working on CLS data will observe the CLS conditions of data use that can be found at the end of the application form, and will not attempt to identify individual cases, or copy any data to other people who may not have signed the appropriate undertaking. All data released from the project to applicants (outside of CLS or a nominated data controller) will be fully anonymised.

### As part of its deliberation about any application, the access committee will apply a ‘Tier’ classification to the data requested (see section 4 for CLS data classification schema).

### If access is granted, this will be administered via a suitable mechanism. For Tiers 1, 2 and 2a, this will generally be via the UK Data Service as set out in Section 4 of this document. However interim arrangements that match the same data security and information governance arrangements as applied by the UK Data Service may occasionally be made, to allow applicant researchers access to this data prior to it becoming available at the UK Data Service.

### In exceptional circumstances where access is granted to Tier 3 data, this would only be permitted subject to suitable confidentiality undertakings and security arrangements being put in place.

### The CLS DAC also considers formal appeals from researchers regarding the Tier level allocated to any CLS data. For further information on how to make such an appeal, please contact [clsfeedback@ucl.ac.uk](mailto:clsfeedback@ucl.ac.uk)

# CLS Data Classification Scheme for Data Sharing

Centre for Longitudinal Studies data is made available to researchers through a number of access mechanisms. The majority of CLS data is available from the **UK Data Service** at:

NCDS: <http://discover.ukdataservice.ac.uk/series/?sn=2000032>

BCS70: <http://discover.ukdataservice.ac.uk/series/?sn=200001>

MCS: <http://discover.ukdataservice.ac.uk/series/?sn=2000031>

Next Steps: <http://discover.ukdataservice.ac.uk/series/?sn=2000030>

The access mechanisms for data available via UKDS depend on the level of risk of disclosure and impact of that disclosure as follows:

* Most CLS data can be accessed via a standard licence known as an 'End User Licence' (EUL). Applications via EUL are authorised directly by the UK Data Service.
* Access to more detailed data, which are potentially disclosive of the identities of individuals, households or organisations is provided via a Special Licence. Applications for data under Special License is administered by the UK Data Service.
* Access to more sensitive or disclosive data is provided via Secure Access which is also administered by the UK Data Service.

Access to biological samples, and some sensitive data derived from biological samples are subject to separate access arrangements. Access to the majority of genotypes generated from NCDS participants is governed by the WTCCC DAC. Access to genotypes linked to other variables (phenotypes), applications for access to DNA, and for new uses of biological samples is via the METADAC (Managing Ethico-social, Technical issues and Administration Data Access Committee).

### CLS Data Classification Scheme

This table illustrates the data classification based on their level of disclosure risk and how the data might be publishable under a particular licence agreement with the UK Data Service at the University of Essex (<http://ukdataservice.ac.uk/>)

|  |  |  |
| --- | --- | --- |
| **Classification** | **Description** | **UK Data Service Data Licence** |
| **PUBLIC** | Publicly available datasets e.g. Edubase, list of schools <http://www.education.gov.uk/edubase/search.xhtml> | n/a |
| **RESTRICTED** - TIER 1 | De-identified survey data e.g. <http://dx.doi.org/10.5255/UKDA-SN-5565-2> | UK Data Service End User Licence |
| **RESTRICTED** - TIER 2a | Data with a medium level of potential disclosure risk or sensitivity, for example counties 1986-2012, e.g. <http://dx.doi.org/10.5255/UKDA-SN-5537-1> | UK Data Service Special Licence |
| **RESTRICTED** - TIER 2 | Data with a high level of potential disclosure risk or sensitivity including: Geography: Output Areas OA, LSOA, MSOA, Local authority, Local Education Authority e.g. <https://discover.ukdataservice.ac.uk/catalogue/?sn=7763> | UK Data Service Secure Access |
| **RESTRICTED** - TIER 3 | Data with a very high level of potential disclosure. Any information which would allow identification of less than 5% of a population of the data item e.g. Postcodes, Date of Birth, School-ID, GP Identifier used for linkage and lookup to other data | n/a – as never deposited via UK Data Service |
| **CONFIDENTIAL** | Individually identifying information e.g. names, address, email, NHS Number, National Insurance Number (NINO), Unique Pupil Number (UPN), Unique Learner Number (ULN), etc. | n/a – as never deposited via UK Data Service |
| **PRIVATE** | Not used. This is used primarily for what can be termed as CLS internal focussed documentation for which there is no benefit or requirement to make it publicly available | n/a |